Selection of Region of Interest in Thermal Images for the Classification of the Human Emotions

Sorin Marius Pavel
Electronics and Telecom. Dept.
Dunarea de Jos University of Galati
Galati, Romania
Sorin.Pavel@ugal.ro

Gabriel Sirbu
Electronics and Telecom. Dept.
Dunarea de Jos University of Galati
Galati, Romania
Gabriel.Sirbu@ugal.ro

Dorel Aiordachioaie
Electronics and Telecom. Dept.
Dunarea de Jos University of Galati
Galati, Romania
Dorel.Aiordachioaie@ieee.com

***Abstract*—The problem of thermal image processing for the classification of human emotions is considered. The work investigates three similarity measures. The first two are based on the statistic correlation of images, i.e., template/reference and candidate/selected images, for the selection of similar regions from a database with thermal images of human faces. The first measure (M1) is based on the Pearson correlation coefficient of two images as sequences, and the second one (M2) is the correlation coefficient extracted from the matrix of correlation of the two images. The last measure uses the structural similarity index (SSIM), alias M3. The reference image is defined by the user for three predefined emotions/states: neutral, sad, and happy. The computer-based simulations reveal the superiority of the correlation-based criterion. Both measures need some constraints to avoid the bad/false/ghost results. Two geometric constraints are proposed based on the decreasing of the searching area to exactly the face area, and – secondly – the left-right symmetry of the selected region. The supplementary constraints impose a two-step selection process: (i) selection of 5-10 image candidates; (ii) extraction based on proposed geometric constraints. The obtained results are satisfactory for the considered thermal database and the set of constraints is general, being valid for other potential similarity measures based on various time-frequency transforms, such as Gabor or wavelet transform.**

Keywords—Thermal images, image processing, feature extraction, similarity measures.

I. Introduction

In image processing, the selection of the region of interest (ROI) is based on measuring the similarity between two images, one as a reference, and the second one, as a working image. The working/processed images come from a database or are cropped from a bigger image. The applications exceed the framework of ROI, going to other important application areas such as image retrieval, object, and face recognition [1]. A major step in the processing process for searching for ROIs is the feature extraction and selection steps, applied to the considered images. The features could be defined and selected from basic histograms, based on the values of the pixels or from more advanced methods like Scale-Invariant Feature Transform (SIFT) or Speeded-Up Robust Features (SURF), which can identify distinctive points in images, and then compared across images. Another important approach is based on the learning machine paradigm, especially the deep learning-based approach. Here, specialized pre-trained neural networks are available like ResNet and VGG, which are used to extract deep features from the processed images, [2]. Finally, once the vector of features is available, the similarity between images can then be computed based on various types of distances, e.g., Euclidean or Manhattan. The previous aspects are valid for visible images. In the case of other types of images, as, e.g., thermal images (from computer vision) or ultrasound images (from SONAR applications), the above methods for ROI selection could fail, mainly because these images are monochrome and with small sample variance of pixel values. So, particular and matched methods perform better. As example, some results are presented now. In [14] a method based on artificial intelligence is used for the ROI extraction for medical use. Reference [15] uses a multi-subject correlation for ROI selection, [16] presents a method for finding a ROI with application in electrical installation, by using local features in describing the region of interest. In the same category of application, [17] uses a model based method, and [18] uses a method based on segmentation and background-subtraction. For thermal image, it is important to adapt the well-known and verified methods applied in the case of visual images, and – eventually - to impose supplementary constraints in the processing of the thermal images. This is the context of the present work, which investigates the problem of ROI in thermal images, corresponding to three states of human emotions, sad, neutral, and happy. The paper is organized into six sections. The next section describes the database of thermal images. In Section 3, the generation of the reference images is considered. In Section 4, the main processing steps for the generation of ROI images are considered. Section 5 presents some results from computer-based experiments. Finally, the conclusion section ends the works and proposes also some further necessary research directions.

II. The Thermal Image Database

The set of the considered thermal images is from the database specially built for thermal images, [3]. Details are available in [4]. There are 161 images for a state, each of 236 x 241 pixels, 24-bit depth, and saved under the bmp file type. An example of such images is presented in Fig.1, for person #20, from a total of 161.

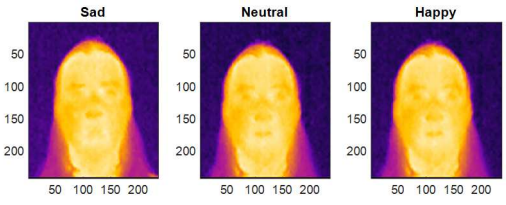

Fig. 1. An example set of thermal images for neutral, sad, and happy.

III. Selection Of The Reference Images

The objective of the section is to describe the selection of the regions of interest (ROI) from thermal images of human faces. Three regions were considered: the region of eyes (E-ROI), the region of the nose (N-ROI), and the region of the mouth (M-ROI). These regions correspond to the anatomical parts of a human face and will be input for the experiments to classify the human emotions from the set of three considered in this work: neutral (N), sad (S), and happy (H). The structure of the method for the ROI's selection is presented in Fig.2.

979-8-3503-4929-0/24/$31.00 ©2024 IEEE

The main blocks are for the computation of the reference image, i.e., the region of interest, in a supervised framework and the blocks for the computation of the ROI, by an automated selection procedure. This includes a feature extraction step, both for reference and for working image.

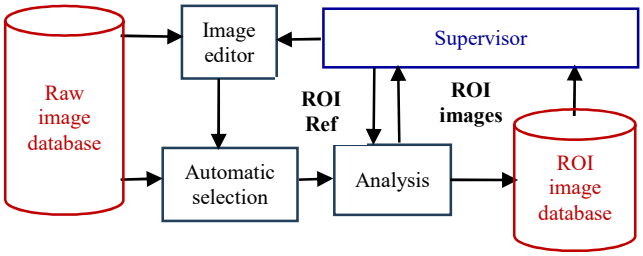

Fig. 2. The structure of the processing method for ROI's selection.

The reference (or template) images could be defined in two frameworks. The first one, which is used in this work, it is based on a subset of the available images, and by selecting manually the ROI. The second way is to use information from virtual faces, as those used in [5], and [6]. The second step for ROI computation is based on searching/exploring, evaluating, selecting, and saving the region of interest. In this work, the exploring means the decomposition of the image in blocks of size equal to the size of the reference image. The selection is made based on the analysis criterion which assesses the similarity between the reference and the processed block.

IV. Similarity Measures

A. The Problem of Similarity

Given two images $\boldsymbol{X}$ and $\boldsymbol{Y}$, a similarity metric as distance $d(\boldsymbol{X}, \boldsymbol{Y})$ is necessary to select the best match between the reference and selected area from a global wider image. Obviously, the similarity function is a metric that satisfy the following four properties [7], [8]:

$$d(\boldsymbol{X},\boldsymbol{Y}) \le d_0;\ \ (Finite\ margin) \tag{1}$$

$$d(\boldsymbol{X},\boldsymbol{Y}) = d_0 \rightarrow \boldsymbol{X} = \boldsymbol{Y}; (Refexitivity) \tag{2}$$

$$d(\boldsymbol{X},\boldsymbol{Y}) = d(\boldsymbol{Y},\boldsymbol{X});\ \ (Symmetry) \tag{3}$$

$$d(\boldsymbol{X},\boldsymbol{Y})d(\boldsymbol{Y},\boldsymbol{Z}) \le \big(d(\boldsymbol{X},\boldsymbol{Y}) + d(\boldsymbol{Y},\boldsymbol{Z})\big)d(\boldsymbol{X},\boldsymbol{Z});\ (Triangle\ ineguality) \tag{4}$$

There are also measures of similarity that do not have all properties of a metric, e.g., those based on joint probability distribution of image intensities, [7]. There are various (standard) similarity measures as: (i) Pearson Correlation Coefficient; (ii) Tanimoto measure; (iii) Stochastic and Deterministic Sign Change; (iv) Minimum ratio; (v) Spearman rank correlation; (vi) Energy of Joint Probability distribution; (vii) mutual information (e.g., Shannon, Renyi, Tsallis). Other measures could be used by considering data transforms, such as Gabor or wavelet transforms, [9], [10]. Selection of the similarity measures needs the validation of some practical constraints as dependence on the size of the reference image, the noise, blurring, and the computational resources including the speed of computation. An important and interesting measure – especially, in the image quality assessment, is the structural similarity index (SSIM), [11], based on three components: luminance, contrast and structure comparison. Thus,

$$L(\boldsymbol{X},\boldsymbol{Y}) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1};\ (luminance) \tag{5}$$

$$C(\boldsymbol{X},\boldsymbol{Y}) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2};\ (Contrast) \tag{6}$$

$$S(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sigma_{XY} + C_3}{\sigma_X + \sigma_X + C_3};\ (Structure) \tag{7}$$

where μ is the sample mean of the image, σ is the sample standard deviation, and σ_{XY} is the sample correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$. The constants involved are used to control the stability of the numerical processing. Obviously, these components are computed by using a local sliding window. The global measure is [12]

$$SSIM(\boldsymbol{X},\boldsymbol{Y}) = L(\boldsymbol{X},\boldsymbol{Y})^{\alpha} \cdot C(\boldsymbol{X},\boldsymbol{Y})^{\beta} \cdot S(\boldsymbol{X},\boldsymbol{Y})^{\gamma} \tag{8}$$

The exponents α, β, and γ define the importance of the components. Common values are $\alpha = \beta = \gamma = 1, C_1 = C_2$ as in [12]. If the elementary processing block has a size less than the processed image, e.g., 8x8 or 16x16, then an average over the SSIM set of blocks is used. The coefficients of Eq. (5-7) are defined as in [12]

$$C_1 = (K_1 A)^2,\ \ C_2 = (K_2 A)^2, C_2 = (K_3 A)^2 \tag{9}$$

where $K_i > 1, i = 1{,}3$ are small constants and A is the range of the pixels (255 for pixel on 8 bit).

B. The used Methods

Three methods were used for selection of ROI. Two are based on correlation, with (i) correlation coefficients (M1) and (ii) correlation matrices (M2). The third one uses the mean of structural similarity index (M3). In all cases, the maximum of the similarity criterion shows the matching coordinates and allow the selection of the desired region for further processing. Thus, let be $\boldsymbol{I}_{ref}$ the reference image and $\boldsymbol{I}_w$ the evaluated image, of equal sizes mxn pixel (the size of the reference image). In the first methods, images are converted in sequences/vectors with elements of the processed images. For two real valued vectors, $\boldsymbol{X}$ and $\boldsymbol{Y}$, each with N elements, the Pearson correlation coefficient measures the linear dependence, and it is defined as

$$\rho(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{N-1}\sum_{i=1}^{N}\frac{X_i - \mu_X}{\sigma_X}\cdot\frac{Y_i - \mu_Y}{\sigma_Y} \tag{10}$$

where μ_X and σ_X are the mean and standard deviation of $\boldsymbol{X}$, respectively, and μ_Y and σ_Y are the mean and standard deviation of $\boldsymbol{Y}$. The conversion of matrices in vectors involves a restriction of equal sizes of the two original/raw arrays (vectors or matrices) and a drawback of losing the 2D information of matrices, which sometimes could be important. The values of the correlation coefficients are in the range [-1, 1]. A small value (close to zero) indicates a poor correlation and values close to one show high correlation. A set K of correlation coefficients is obtained as

$$K = \{k_j, j = 1,2, \dots, n\} = corr(\boldsymbol{I_{ref}}, \boldsymbol{I_{wj}}) \quad (11)$$

To keep the 2D information and not have the restriction of size, the 2D cross-correlation is used. This is method M2. For two real value arbitrary size matrices $\boldsymbol{X}$ (m x n) and $\boldsymbol{Y}$ (p x q) the 2-D cross-correlation is a matrix, $\boldsymbol{C}$, of size m+p–1 by n+q–1. Its elements are given by [13]

$$C(k,l) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} X(i,j)Y(i-k, j-l), \quad -(p-1) \le k \le m-1, \quad -(q-1) \le l \le n-1 \quad (12)$$

The output matrix, $C(k, l)$, has negative and positive row and column indices. A negative row index corresponds to an upward shift of the rows of $\boldsymbol{Y}$. A negative column index corresponds to a leftward shift of the columns of $\boldsymbol{Y}$. A positive row index corresponds to a downward shift of the rows of Y. A positive column index corresponds to a rightward shift of the columns of $\boldsymbol{Y}$, as described by [13]. The maximum coefficient of the matrix $\boldsymbol{C}$ is selected and considered as the value of the correlation between matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$. The low values indicate that the two matrices are not similar.

Let be S the set of the (working) images

$$S = \{\boldsymbol{I_{wj}}, j = 1,2, \dots, n\} \quad (13)$$

The size of each image is the same as the reference image. The selected image is $\boldsymbol{I_s}$, which has the highest correlation coefficient from the set K:

$$s = \underset{j}{argmax}\{K_j, j = 1,2, \dots, n\} \quad (14)$$

The selection is implemented for all three types of emotions, and thus – finally – a set of selected images, one for each considered state (neutral, sad, and happy), is obtained. The third method (M3) uses the SSIM approach (presented in the previous section) with values.

$$K = [1, 1, 100] * 255 \quad (15)$$

V. Experimental Results

The reference images (one for each state) are built by randomly selecting 10% from each set of images, i.e., 16 images, and by computing an average of the pixel's values and the size equal with the minimum size from the processed set. Thus, the reference images have 20 to 25x60 pixel. The selection of the regions is made manually based on visual inspection and selection. The results are presented in Fig. 3. There are three reference images for each state (happy, normal, and sad). Table I presents the numeric results obtained for ROI selection based on the three evaluated methods. The performance is defined as the correct selection as the ratio between the correct selection and total processed images. The values in bold and yellow background are for the best values from a set of three. The best results are obtained with M1 (correlation coefficient based) in six cases, followed by M3 in two cases, and M2 in one case.

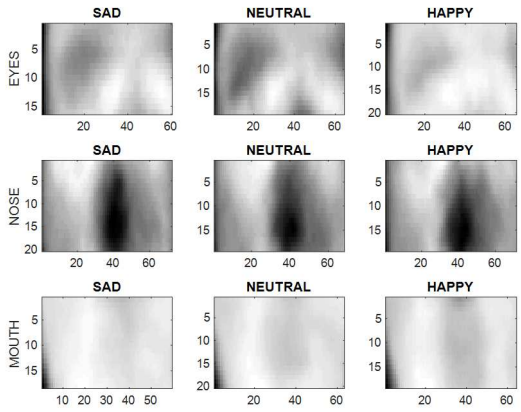

Fig. 3. The reference images for the three states and three regions.

Table II presents an estimation of the complexity of the considered method from the point of computation time on a Windows 11 machine with i7 – 1.70 GHz, 64-bit operation system and Matlab 2018 simulation software. The smallest complexity is obtained with M1 (based on correlation coefficient).

TABLE I – PERFORMACE OF THE ROIS SELECTION

Reference / template	State	Method		
		Corr-coef	*Corr-mat*	*SSIM*
Eyes	Happy	18.01	59.62	**67.70**
Nose		**72.67**	50.93	49.69
Mouth		**54.65**	17.39	31.05
Average		***48.44***	*42.64*	***48.48***
Eyes	Neutral	29.19	**52.17**	42.85
Nose		**72.04**	26.08	19.87
Mouth		**48.44**	18.01	21.32
Average		***49.89***	*32.08*	*28.01*
Eyes	Sad	32.30	47.82	**50.93**
Nose		**63.35**	19.25	27.32
Mouth		**53.41**	19.88	20.49
Average		***49.68***	*28.98*	*32.91*

TABLE II – COMPUTATION TIME [S]

Reference / template	State	Method		
		Corr-coef	*Corr-mat*	*SSIM*
Average time [s]		**5.87**	166.66	230

Fig. 4 presents an example of the selection process for M1, region nose, and state happy. The selected image is highlighted in red contour. There are 13 good and 3 wrong selections. A comparison between M1 and M2 is presented in Fig. 5, for nose selection. On the upper side, the values of the selection criteria are presented. On the bottom side, the reference and the selected images are shown. There is a difficulty to select the maximum peak of the cross-correlation vector or matrices because there are relatively many maximum points. This means that the evaluated regions are quite similar, and this could generate wrong results related to the right selection of ROI. This general remark is valid also for M3, as in Fig. 6, where a wrong selection is obtained (mouth instead of eyes region).

The preliminary analysis of the obtained results reveals some shortcomings, in the sense that there are false ROIs, 5-10 % from the size of the actual database. This is generated when some bad images are processed, e.g., distinct size and orientation of the face. The next step of the research will try to decrease this, by implementing features from artificial /synthetic ROI or by correcting the inputs.

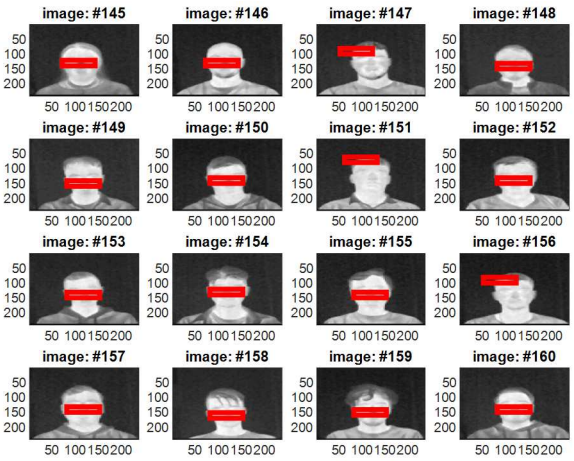

Fig. 4. Results in the selection of ROI; nose region, happy state.

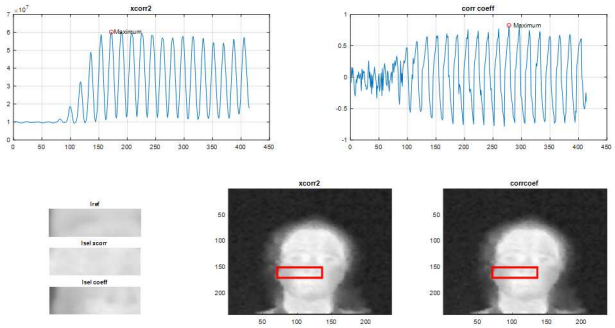

Fig. 5. Results of selection of ROI; nose region, happy state.

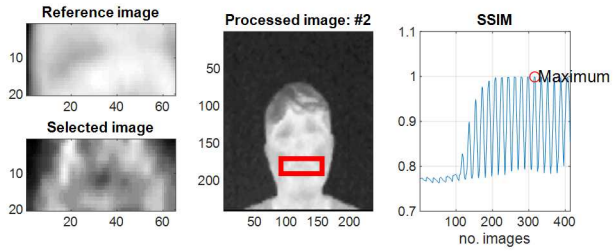

Fig. 6. Results in using SSIM criterion; eyes region, happy state.

Because of the relatively high number of bad results in selecting ROIs, the previous similarity measures are modified by conditioning by two geometric constraints as: (i) the searching area is restricted to the area of a human face, and not the entire image; (ii) the selected ROI should satisfy a left-right symmetry, in the sense that the first half of the selected image (from de left side) should be close to the second half from the right side. The previous constraints impose a two-step selection process. In the first step, based on the similarity measure, a set of 5-10 ROIs are selected. In the second step, the ROI which satisfies all the constraints will be proposed as the winner of the search and selection process.

CONCLUSION

The objective of the work was to evaluate three methods of image similarity for the selection of the sub-images or blocks, based on a template/reference input. Even the thermal images are simple, content plus intensities, i.e., they are monochrome images, these are difficult to process, in the sense of applying the well-known and mature similarity methods verified already on visible images. The content of the monochrome thermal images does not allow us to find the ROIs of small sizes, i.e., less than 50 by 50. The reason is the low variance in the human face region. To avoid this, and for the objective of detection and classification of the emotions, the interest regions should be extended to combinations of two, i.e., eyes plus nose and, second, mouth plus nose. The right selection of ROIs on human faces needs additional constraints set related to the searching space, i.e., the searching space should be limited by the face area and should satisfy the left – right symmetry property. These constraints imposed a two-step selection process. All these measures will be considered in the next research steps. Some time-frequency transforms, as Gabor or wavelet transform, will be considered also.

REFERENCES

[1] Medium Ltd, 2024. Best Image Similarity Search APIs in 2023, Accesed on 27.02.2024, URL: https://medium.com/@edenai/best-image-similarity-search-apis-in-2023-45fe37b41a24.

[2] A. Asperti and D. Filippini, "Deep Learning for Head Pose Estimation: A Survey," SN COMPUT. SCI. 4, 349, 2023.

[3] M.S.Pavel and D. Aiordachioaie, Thermal Image database, Signals and Information Processing Laboratory of CCETIC, from the University „Dunarea de Jos" of Galati), 2023. Accessed link: http://www.etc.ugal.ro/ccetic/eng/database2.html

[4] M.S.Pavel and D. Aiordachioaie, "Processing A Database With Thermal Images For The Classification Of Emotional States," The Bulletin of the Polytechnic Institute of Iași, Gheorghe Asachi Technical University of Iași, Romania, Electrical Engineering, Power Engineering, and Electronic section, 2024 (*in review*).

[5] A. M. Alattar and S. A. Rajala, "Facial features localization in front view head and shoulders images,", IEEE Int. Conf. on Acoustics, Speech, and SP, Phoenix, AZ, USA, 1999, vol. 6, pp. 3557-3560.

[6] J.S.Sheu, T.S. Hsieh, H.N. Shou, "Automatic Generation of Facial Expression Using Triangular Geometric Deformation," Journal of Applied Research and Techn, Vol. 12, Issue 6, 2014, pp. 1115-1130.

[7] A.A. Goshtasby, Image Registration, Springer, 2012.

[8] A. Ciobanu, T. Barbu and C. Niţă, "Novel image similarity metric for evaluating denoising and restoration techniques," 2017 E-Health and Bioengineering Conference (EHB), Romania, 2017, pp. 470-473.

[9] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik and M. K. Markey, "Complex Wavelet Structural Similarity: A New Image Similarity Index," in IEEE Trans. Image Proc., vol. 18 (11), pp. 2385-2401, 2009.

[10] D. Vukadinovic and M. Pantic, "Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers," The IEEE Int. Conf. on Syst., Man and Cybernetics, Waikoloa, Hawaii, 2005.

[11] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in The Handbook of Video Databases: Design and Applications, B. Furht and O. Marques, Eds. CRC Press, 2003.

[12] M. A. Hassan and M. S. Bashraheel, "Color-based structural similarity image quality assessment," 2017 8th International Conference on Information Technology (ICIT), Amman, Jordan, 2017, pp. 691-696.

[13] Mathworks, Accesed at 30.01.2024. URL: https://www.mathworks.com/help/signal/ref/xcorr2.html.

[14] L. C. Mendes, E. O. Rodrigues, S. C. Izidoro, A. Conci and P. Liatsis, "ROI Extraction in Thermographic Breast Images Using Genetic Algorithms," 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 2020, pp. 111-115.

[15] K. Hong, "Classification of Emotional Stress and Physical Stress Using Electro-Optical Imaging Technology," 2nd Int. Conf. on Networking Systems of AI (INSAI), Shanghai, China, 2022, pp. 92-97.

[16] M. S. Jadin, K. H. Ghazali, S. Taib and N. Huda, "Finding ROIs in infrared image of electrical installation for qualitative thermal condition evaluation," IEEE Int. Conf. on Control System, Computing and Eng., Penang, Malaysia, 2012, pp. 244-249.

[17] E. K. Lee, H. Viswanathan and D. Pompili, "Model-Based Thermal Anomaly Detection in Cloud Datacenters Using Thermal Imaging," in IEEE Transactions on Cloud Computing, vol. 6 (2), 2018, pp. 330-343.

[18] J. W. Davis and V. Sharma, "Robust detection of people in thermal imagery," Proceedings of the 17th Int. Conf. on Pattern Recognition, ICPR 2004., Cambridge, UK, vol.4, pp. 713-716.